

Zitate verkaufen? Bemerkungen zur Ethik des Geschäfts mit Metadaten

Herbert Hrachovec

Zusammenfassung

Veröffentlichungen hinterlassen eine bis vor Kurzem unrealisierbare Spur im Internet. Der Wissenschaftsbetrieb ist davon zunehmend betroffen. Mächtige Firmen sammeln solche Informationen und werten sie für ihre Zwecke aus. Als Beispiele werden Google Scholar und academia.edu vorgestellt. Ernsthafte Bedenken gegen den sich abzeichnenden verstärkten Gebrauch solcher Daten sind angezeigt.

Abstract

Selling Citations. Remarks on the Ethics of trading metadata

The trace a publication leaves on the internet gives rise to an unprecedented amount of information about its use. Academic life is affected by the ability of powerful companies to collect and exploit such data. Google scholar and academia.edu are presented as exemplary cases. It is argued that serious concerns can be raised against the way these technical advances are increasingly being used.

Die Veröffentlichung eines Buches ist in der Regel nicht mit ethischen Problemen verbunden. Ein Verlag nimmt es ins Programm, bewirbt und verbreitet es. An diesem Ablauf ändert das Internet prinzipiell nichts. Die einfachere und schnellere Distribution des Produktes lässt den Kern der Sache unberührt. Auch die Option, Bücher elektronisch anzubieten, bleibt im gewohnten Rahmen – abgesehen davon, dass sie in diesem „Aggregatzustand“ leichter unerlaubt zu verteilen sind. Anders verhält es sich mit der Wirkungsgeschichte der Publikation. Die bibliographische Spur einer Veröffentlichung hat sich mit dem World Wide Web verlängert und verbreitert. Sie besteht nicht mehr bloß darin, dass sich Einträge in Bibliotheken finden, die das Buch erworben haben.

Es liegt schon lange zurück, aber es gab Sammlungen, in denen Büchern ein Entlehnzettel eingeklebt wurde, der Details ihrer Verwendung registrierte. So konnte man einen Eindruck von der Benutzungsgeschichte des jeweiligen Exemplars gewinnen. Diese simplen Aufzeichnungen werden mittlerweile durch ein gewaltiges Datennetz überboten. Die Suchergebnisse, die Google liefert, geben Aufschluss über die Popularität einer Publikation, und der Verkaufsrang bei Amazon erlaubt eine Kalkulation ihres Marktstatus. In

dieser neuen Umgebung entstehen unerwartete Fragen hinsichtlich einiger Implikationen des technisch transformierten Arrangements. Ein Beispiel ist die Möglichkeit, faktische Suchvorgänge auf einer bestimmten Webseite durch Zähler zu dokumentieren, oder, wie bei Google, als wahrscheinliche Phrasen bereits beim Eintippen zu antizipieren. Solche Informationen sind leicht manipulierbar, und das ist in einigen Fällen auch nicht verwerflich. Unter welchen Umständen ist es berechtigt, die Ergebnisse zu verfälschen? Terror- und Pornoseiten sind häufig angeführte Beispiele. Von solchen Fällen ist nur ein kurzer Schritt zum allgemeinen Problem, welche Verantwortung eine Instanz trägt, die Daten über den massenhaften Umgang mit Angeboten im Internet verwaltet.

Im Folgenden wird es um Entwicklungen gehen, die es gestatten, den globalen Umgang mit akademischen Veröffentlichungen zu verfolgen und auszuwerten. Die Aufmerksamkeit wird auf den Umstand gelenkt, dass die Akkumulation im Prinzip öffentlicher Daten in zentralen, zugangsbeschränkten Sammelstellen spezifische ethische Fragen aufwirft. Das Problem ist nicht neu. Einerseits erscheint es unumgänglich, dass eine Gesundheitsbehörde Statistiken der Todesfälle in Krankenhäusern erstellt. Andererseits sind die Bedenken gegen eine freie Veröffentlichung dieser Ergebnisse zumindest verständlich. Ihre öffentliche Rezeption kann ein unzulässig vergrößertes Bild geben und zu Kurzschlussreaktionen führen. Wir werden die digitale Spur verfolgen, welche wissenschaftliche Bücher und Zeitschriftenartikel hinterlassen. Im Zeitalter der weitverbreiteten, instantan abfragbaren Information gewinnt sie immer stärkere Bedeutung in der Eigen- und Fremdwahrnehmung der Profession.

Zwei Aggregationsformen bibliometrischer Information sind von speziellem Interesse. Sie markieren unterschiedliche Positionen im Kräftespiel zwischen Öffentlichkeit, Datenauswertung, kommerziellem Interesse und verantwortlichem Umgang mit präsumtiven Resultaten. Der naheliegende Einstieg ins Thema ist „Google Scholar“¹, die Erfassung der Metadaten wissenschaftlicher Veröffentlichungen, inklusive der zwischen ihnen bestehenden Verweisstrukturen im von der Firma abgedeckten Suchraum. Alternativ zu diesem generalisierenden Verfahren haben sich Angebote entwickelt, die für eingeschränkte Personenkreise die Möglichkeit eines selbstverwalteten Datenregimes anbieten. Beispiele für solche Forschungsclubs sind etwa <http://www.researchgate.net>, <https://www.linkedin.com/academia.edu>. In der Datenwolke des Internet haben sich konkurrierende Praktiken des Umgangs mit der Wirkungsgeschichte von Forschungsergebnissen herausgebildet. Sie sind nicht wertneutral, sondern implizieren unterschiedliche Auffassungen über Privatheit und Öffentlichkeit kreativer Produktion und die Auswirkung ihrer Zugänglichkeit auf das Gemeinwohl.

¹ <http://scholar.google.com>

Google Scholar

Das Ausmaß, in dem Google mit seinen diversen Diensten Charakteristika der Benutzerinnen gewinnen kann, ist beängstigend [Tene, 2008]. Zur vergleichsweise konventionellen Strategie der auf Zielgruppen abgestellten Werbung ist durch die Mobilkommunikation die örtliche Erfassung von Individuen gekommen, welche für den Komfort persönlicher ad-hoc Services die Erfassung ihrer Tagesabläufe einhandeln. Um diesen Themenkomplex geht es hier nicht. Der Fokus liegt auf der Verarbeitung von Daten aus der Wissenschaftspraxis. Die Basis des Suchdienstes bilden Informationen aus verschiedenen Quellen. Einerseits solche, die der Buchhandel von altersher angeboten hat: Verlagsbroschüren und Verkaufslisten, andererseits solche aus dem Fundus der Suchmaschine: Dissertationen, Preprints, technische Expertisen und Arbeiten, die sich auf den persönlichen Webseiten des akademischen Personals von Universitäten finden. Die periodisch wiederkehrende globale Datenerfassung fügt einen Aspekt ohne Präzedenz hinzu. Im Netz entfaltet sich ein Verlaufsbild der Nutzung dieser Publikationen.

Bestseller, von denen „alle Welt spricht“, drängen sich auch in Google auf. Wie sieht es mit wissenschaftlicher Produktion aus, die solchen Gesetzmäßigkeiten nur in den seltensten Fällen folgt? Google Scholar hat ein aus der prä-digitalen Zeit stammendes Paradigma zur Abbildung der Breitenwirkung der Fachliteratur an die Verhältnisse des WWW adaptiert.² Dieser Dienst dokumentiert Veröffentlichungen im Wissenschaftsbereich und Zitate, in denen sie sich aufeinander beziehen.³ Der Nutzen solcher Auflistungen steht außer Frage. Worum geht es in der Wissenschaft sonst, als sich vernehmbar zu machen und in Diskussionszusammenhänge einzuschalten?

Die 2004 eingerichtete Wissenschafts-Datenbank basiert auf einigen konstitutiven Entscheidungen, welche die neuartige Situation erforderlich macht. [Jacsó, 2005], [Noruzi, 2007], [Harzing and Van der Wal, 2007] Google verfügt über einen enormen Datenvorrat und beherrscht den Zugang zu detaillierten Informationen über das

2 „Google Scholar ordnet Suchergebnisse nach Relevanz. So wie bei der Webseitensuche mit Google werden die nützlichsten Verweise oben auf der Seite angezeigt.“

<http://scholar.google.at/intl/de/scholar/about.html> 4.1.2014

3 „Die Ranking-Technologie von Google berücksichtigt den vollständigen Text eines Artikels, den Autor, wo der Artikel veröffentlicht wurde und wie oft der Text in der wissenschaftlichen Literatur zitiert wurde.“ a.a.O.

dazugehörige Nutzerverhalten. Das Unternehmen stellt sich die Aufgabe, die interne Struktur der Teilmenge „Forschungsliteratur“ adäquat wiederzugeben. Datenmaterial in diesem Umfang ist bisher nur unter der Ägide staatlicher Institutionen angehäuft worden. Es unterlag, zumindest im Prinzip, politischer Kontrolle. Das österreichische Universitätsgesetz 2002 schreibt den Hochschulen die Erstellung einer „Wissensbilanz“ vor⁴, die alle relevanten Forschungsleistungen bottom up dokumentieren soll. Das Ansinnen verblasst hinter den Möglichkeiten des Weltkonzerns.

Auf welcher Datengrundlage beruht das Angebot? Wie aktuell wird der Bestand gehalten? Nach welcher Methode werden daraus die relevanten Faktoren extrahiert? Wodurch ist seine Kohärenz gewährleistet und welche Korrekturmechanismen sind vorgesehen? Die Antworten sind nur teilweise öffentlich bekannt. [Harzing and van der Wal, 2008], [Bowen, 2014] Google greift sicherlich auf seinen Fundus an Suchabfragen zurück und bedient sich, wie gesagt, des Materials von Verlagen, Bibliotheken, elektronischen Repositorien und akademischen online verfügbaren Veröffentlichungen. Eine Aufstellung der Quellen fehlt allerdings. Es ist auch unbekannt, in welchen Abständen sie erneuert werden und wie, beziehungsweise wann die Überprüfung auf inkorrekte Einträge stattfindet. Man weiß, dass Google sich zur Datenanalyse nicht auf die in vielen Fällen verfügbaren bibliographischen Metadaten stützt [Jacsó, 2009], sondern einen eigenen Algorithmus zur Aufschlüsselung der semantischen Struktur einsetzt, welche die „crawlers“ auf Webseiten finden. [Mikki, 2009], [Gray, 2012] Google Scholar ist das frei zugängliche Angebot eines Unternehmens mit mächtigen Ressourcen, das bloß beschränkt Einblick in die Interna seiner Betriebsführung bietet.

Die Vorgänge, die auf die angedeutete Weise ausgewertet werden, finden im öffentlichen Raum statt und sind traditionell als Allgemeingut betrachtet worden. Bibliothekskataloge standen zum geteilten Gebrauch aller Interessentinnen zur Verfügung, und die Zitarnachweise im wissenschaftlichen Apparat wurden nicht durch Datenbanken im Besitz kommerzieller Anbieter überboten.⁵ An „Google Scholar“ wird der technische Umbruch deutlich, der aus einem Separatraum für Gelehrte die Unterabteilung einer weltumgreifenden Datenwolke unter Kontrolle weniger Akteure

⁴ http://www.reko.ac.at/modules/download.php?key=2833_DE_O&cs=970C

⁵ Auf die Rolle des „Web of Science von Thomson Reuters“ wäre in diesem Zusammenhang gesondert einzugehen.
<http://thomsonreuters.com/web-of-science/>

gemacht hat. Die eben notierten Designvorgaben des Unternehmens sind auf diese Weise auch Weichenstellungen in der Konfrontation mit dem Problem, inwiefern ein Gemeinwesen sein Wissen und seine Aktivitäten zur Privatisierung freigeben kann und soll. Aus der Auseinandersetzung über das menschliche Genom oder die Patentierung von Wirkstoffen indigener Medizin ist der Typus der Schwierigkeit bekannt. An „Google Scholar“ manifestiert er sich für bibliographische Kontexte.

Hier kann es nicht um einen systematischen Beitrag zur Thematik gehen. Eine Besonderheit des angesprochenen Spezialfalls ist aber jedenfalls zu unterstreichen. Google vereinnahmt seine Resultate *gerade nicht*. Die Suchfunktion stellt Ergebnisse zur Verfügung, die einen von der Firma generierten Mehrwert allgemein zugänglich machen. Einwände betreffend Qualität und Validität können bei einem frei angebotenen Produkt (das seinerseits auf frei gewählter Informationsweitergabe beruht) nur beschränkt greifen. Das Konfliktpotenzial liegt anderswo. Die konkurrenzlose, kostenfreie Erschließung eines Suchraumes ist derart attraktiv, dass Nachteile auf weite Strecken hingenommen werden. Péter Jacsó hat darauf hingewiesen, dass durch den im Eigenbau konstruierten Algorithmus eine große Zahl von Autorinnen als „Password“ „Introduction“ und Ähnliches angegeben werden, womit ihr echter Name aus dem Eintrag verschwindet.

Google's algorithms create phantom authors for millions of papers. They derive false names from options listed on the search menu, such as P Login (for Please Login).

Very often, the real authors are relegated to ghost authors deprived of their authorship along with publication and citation counts. In the scholarly world, this is critical, as the mantra publish or perish changes to publish, get cited, or perish. [Jacsó, 2009], vgl. [Kraft, 2009]

Solche Einwände sind amüsant, lassen sich aber durch ein Update leicht beheben und ändern nichts daran, dass Google Scholar für den gewöhnlichen Gebrauch (wenn man auf Englisch publiziert und einen Namen ohne Umlaute besitzt) befriedigende Resultate liefert. Auch der Nachweis, dass die von Google Scholar ausgewiesenen Zitationsdaten ziemlich einfach zu manipulieren sind [Bar-Ilan, 2007], [Delgado López-Cózar et al., 2012], [Anderson, 2013], hat kein breites Echo gefunden. Die Defizite, welche auf die Entscheidungen zur Implementierung dieses Dienstes zurückgehen, werden weithin als beiläufige Irritationen eines praktisch hilfreichen Angebotes akzeptiert.

Bedenklich ist nicht Googles Experimentalanordnung im großen Stil, sondern die Anwendung dieses bibliometrischen Programms als

Instrument zum Management der Forschungssteuerung. Sie basiert auf Voraussetzungen, die nicht von Google stammen, sondern von den Verantwortlichen in Politik und akademischem Management vorgegeben werden. Meteorologische Vorhersagen dürfen unpräzise sein, es sei denn, die Umstände verlangen eine exakte Prognose. In einem solchen Fall trifft die Verantwortung die Personen, die sich dieses Mittels bedienen. Ein ähnliches Dilemma ergibt sich durch den zunehmend verbreiteten Einsatz bibliometrischer Daten, darunter auch der Ergebnisse von Google Scholar, zur Beurteilung wissenschaftlicher Reputation und Karriereverläufe. Das Wissen um die mögliche Fehlerhaftigkeit der Kennzahlen, von denen zunehmend Entscheidungen über Personen abhängig gemacht werden, muss in den Kalkül eingehen. Die Spuren einer Veröffentlichung im Zeitablauf zu verfolgen ist nicht, wie früher, auf explizite Zitate in der Fachliteratur beschränkt. Sie tauchen in Blogeinträgen und Forumdiskussionen auf, werden gesucht und über Twitter und Facebook verbreitet. Google erfasst einen erheblichen Teil dieser Abläufe, doch man weiß nicht, nach welchen Filtern. „Citation analysis needs an in-depth transformation. Current systems have been long criticized due to shortcomings such as lack of coverage of publications and low accuracy of the citation data. Surprisingly, incomplete or incorrect data are used to make important decisions about researchers’ careers.“ [Cerdá et al., 2009]

Die Tabelle einer Fußball Liga oder die Ergebnisse eines Songkontests werden selbstverständlich veröffentlicht. Auch die Publikationsliste einer Wissenschaftlerin kann kein Geheimnis bleiben. Aber wie verhält es sich mit jenen Daten, die eine fremde Instanz, nämlich ein Internet-Provider, im Anschluss an ihre Einträge zusammenstellt? Sie werden systematisch registriert und ausgewertet, ohne dass der Betroffenen ein kodifiziertes Mitspracherecht zukommt. Als einfaches Beispiel kann man sich den Funktionsumfang eines Servers, der ein open access Datenarchiv verwaltet, vor Augen halten. Die Einreichungen erfolgten sicherlich mit der Absicht, die Arbeiten einer möglichst großen Öffentlichkeit zur Verfügung zu stellen. Daraus folgt aber nicht, dass alle Auswirkungen dieses Entschlusses überblickt und akzeptiert werden. Jeder Zugriff auf die Webseiten und jeder Download eines Textes wird in den Logfiles des Servers detailliert verzeichnet (Datum, Uhrzeit, Herkunftsort). Unversehens wird aus einem freien Angebot im Web Material für statistische Nachbearbeitung. Ein wissenschaftlicher Beitrag findet sich als Bestandteil globaler informatischer Transaktionen.

Zur Verdeutlichung einige Eckdaten eines Buches in digitaler Form, das 2013 auf <http://sammelpunkt.philo.at> verfügbar gemacht wurde. Es zählt am 29.12.2013 248 Downloads, davon 141

aus Deutschland, 32 aus Österreich und 11 aus den USA. So weist es eine Analysesoftware für den Webserver aus und damit beginnen die Schwierigkeiten. Die Zählung durch ein alternatives Programm ergibt für den betroffenen Posten die Zahlen 285, 182, 40 und 8. Beide Statistiken greifen auf *dasselbe Protokoll* der digitalen Transfers zurück, das jede einzelne Nachfrage registriert. Über diese Pauschalangaben hinaus ist nachvollziehbar, dass das Buch sofort nach der Veröffentlichung von Facebook aus abgerufen wurde und in den ersten beiden Tagen auch aus der Universität Trier, sowie über einen deutschen und einen US-amerikanischen Provider. Die präzise Uhrzeit gestattet im Prinzip, die Maschine, auf welcher die Nachfrage abgesetzt wurde, festzustellen. Das gilt in gleicher Weise für alle 1755 Einträge, die sich auf dieses Fundstück beziehen, nur dass die Auswertungen und Zuordnungen der Herkunftsadressen, welche die Diskrepanzen in der Analyse zeigen, offenbar nicht zu eindeutigen Ergebnissen führt.

Alexander Dilger und Harry Müller weisen in ihrer Studie über Publikationsprofile in der deutschen Betriebswirtschaftslehre bloß die zehn meistzitierten Personen namentlich aus. [Dilger and Harry Müller, 2012, S. 9] Rangreihen sind umstritten, je weiter sich die Domäne von eindeutig quantitativen Befunden entfernen. Zur Ethik bibliometrischer Forschungsdesigns gehört die Einschätzung, welche Effekte die Mobilisierung der seit Kurzem zur Verfügung stehenden Datenmengen auf den Wissenschaftsbetrieb und seine Trägerinnen hat. Der Spitzenplatz in einer Liste ist unwillkürlich auch ein Werturteil. Es ist zwar richtig, dass in einem methodisch sauberen Verfahren die Arithmetik von der Nutzenanwendung getrennt wird, dennoch sind Formeln, welche zu handfesten Ergebnissen führen sollen, nur künstlich wertfrei. Wenn man die starken Ungleichheiten bedenkt, die zwischen Sprachen, Wissenschaftskulturen, finanziellem Aufwand und interkulturellem Prestigedenken bestehen, ist die methodische Vorgabe der Vergleichbarkeit keine Selbstverständlichkeit, sondern ein gezielter Eingriff in die Untersuchungsumgebung, also eine verantwortliche Entscheidung.⁶ Und umso mehr gilt das für Schlüsse, die auf einer derartigen Grundlage gezogen werden.

6 Dazu als Vignette: Die von Google Scholar angegebenen Kennzahlen des meistzitierten englisch- respektive deutschsprachigen Journals ergeben einen h5-Index von 349 („Nature“) bzw. 22 („Deutsches Arzteblatt-Arztliche Mitteilungen-Ausgabe B“). http://scholar.google.at/citations?view_op=top_venues&hl=de&vq=en

academia.edu

Die bisher angeführten Entwicklungen betreffen den Wissenschaftsbetrieb insgesamt und wirken gleichsam hinter dem Rücken der Akteure. Im Kontrast dazu entsteht durch Innovationen im „social web“ eine individuellere Konstellation. Strategien zur aktiven, freien Verbreitung der eigenen Arbeitsergebnisse passen gut zum wissenschaftlichen Ethos. Netzgemeinschaften bieten die Möglichkeit, persönliche Profile anzulegen und Zusammenhänge im Fachdiskurs herzustellen. Sie gehen über die Bibliometrik hinaus und importieren einen repräsentativen Teil der professionellen Tätigkeit in einem neuen Medium. Die Vorteile solcher vernetzter Interaktionen sind ausführlich dargestellt worden. [Benkler, 2006], [Castells, 2011] Allerdings hat sich in den Erfahrungen mit Facebook, Twitter und Google+ auch herausgestellt, dass diese auf den ersten Blick kleinteiligere Angebote ihrerseits Datensammlungen in großem Maßstab sind. Mit dem Kontrast zu den Beobachtungen des vorangegangenen Abschnitts ist es nicht weit her.

Die Situation im akademischen Bereich ist wiederum vor dem Hintergrund der allgemeineren Innovationen zu sehen. Durch „Freunde“ wird ein Bezugsrahmen abgesteckt, der die bei der Anmeldung der jeweiligen Person erhobenen Eckdaten anreichert. In diesem Zusammenhang sind Präferenzen und Vorlieben („likes“) wertvolle Parameter zur Erfassung bestimmter Milieus. Maßgeschneiderte Serviceprogramme („Apps“) für Nachrichtendienste, Musik- und Videobeschaffung, Sport, Konsum und Unterhaltungsangebote fragen Informationen über das tägliche Leben ab und integrieren es in großflächige Muster. Die Unterlagen für gezielte Werbekampagnen liefern die Betroffenen selbst. Facebook seinerseits verkauft z.B. den TV-Stationen Informationen über Alter, Geschlecht und Wohnort jener Teilnehmerinnen, die sich über ihre Sendungen äußern. [Bénilde, 2013] Man kann erwidern, dass solche monetären Motive im Wissenschaftsbereich nicht direkt greifen, doch dabei wird ein wichtiger Umstand übersehen: Die Gesetze des Datenverkehrs und seiner inhaltlichen Aufschlüsselung machen keinen Unterschied zwischen Konsumverhalten und Forschungsergebnissen. Sie gelten für Erdnussbutter ebenso wie für Expertenwissen. Und in einer durch die Ökonomie des Informationszeitalters gekennzeichneten Lage versprechen statistische Auswertungen über das Sozialverhalten von Wissenschaftlerinnen einen besonderen Mehrwert.

Zu beobachten ist das beispielhaft an der Geschäftspolitik des Portals academia.edu, dessen Gründer, Richard Price, den Zweck des Unternehmens folgendermaßen beschreibt:

The goal is to have every single science PDF ever written available for free on the Internet and to build a network of scientists interacting with those papers that will change the face of peer review. [Shankland, 2013]

Dieses ambitionierte Ziel deckt sich prima vista mit den Manifesten der open access-Bewegung⁷ und verspricht den Aufbau einer wissenschaftlichen Gemeinschaft jenseits der Zitatenzählung. Das Ansinnen ist ambitiös, erweist sich aber bei näherem Hinsehen als illusorisch. academia.edu bietet sich als Anlaufstelle für die Erstellung persönlicher Forschungsprofile an. Auf ihr können die Aktivitäten, anders als in den gewöhnlichen institutionellen Rastern, eigenbestimmt zusammengestellt werden. academia.edu hilft dabei, bereits verfügbare Veröffentlichungen an dieser Stelle zu bündeln und lädt dazu ein, zusätzliches Material hochzuladen. Von einer vollständigen Erfassung der wissenschaftlichen Produktion („every single science PDF ever written“) ist das meilenweit entfernt.

Die Mitgliedschaft von Forscherinnen bei academia.edu müsste flächendeckend sein, und die Limitationen, die durch das Verlagswesen und seine hochdifferenzierten Verwertungsrechte eingemahnt werden, müssten ganz unerwartet wegfallen. Die FAQ zum Portal weist darauf hin, dass die Aufstellung auf Sherpa/RoMEO bei 90% der wissenschaftlichen Journale die Bereitschaft feststellt, Beiträge in irgendeiner Form zur offenen Archivierung freizugeben.⁸ Doch die Details sind mühsam⁹ und Buchpublikationen sind in dieser Aufstellung nicht berücksichtigt.

⁷ http://open-access.net/at_de/startseite/

⁸ <https://www.academia.edu/FAQ>,
<http://www.sherpa.ac.uk/romeo/>

⁹ Hier die differenzierte Formulierung auf der Originalseite: „This chart shows that a remarkable 94% of journals allow archiving of peer-reviewed articles after any embargo period has expired and any additional restrictions have been complied with. Indeed, for nearly a quarter of journals, the publisher’s version/PDF itself can be archived. Just 1% of journals only permit the pre-peer review submitted version to be archived. This leaves only 5% of journals that do not permit self-archiving of some form or another.“
<http://romeo.jiscinvolve.org/wp/2011/11/24/60-of-journals-allow-immediate-archiving-of-peer-reviewed-articles-but-it-gets-much-much-better/> 4.1.2014

Die Diskrepanz zwischen der maximalistischen Ansage und der realen Reichweite dieser „social site“ erklärt sich durch handfeste Interessen, die sich hinter dem Angebot einer neu entwickelten globalen Kollegenschaft im Netz verbergen. Die Firma ist durch Risikokapital finanziert.¹⁰ Was erweckt das Interesse der Kapitalgeber an solchen Beteiligungen? An diesem Punkt schlägt die Logik und Ökonomie der Internet-Giganten auf den vergleichsweise eingeschränkten Bereich der Wissensproduktion durch. Die Mitgliedschaft von Einzelpersonen in einem digitalen Club ist ein Vehikel zur Aggregation von Daten, die nach Richard Price zur Entdeckung von Forschungstrends z.B. in der Pharmakologie und zur Verbesserung der Personalpolitik großer Unternehmen ausgewertet werden können:

„When we’ve built what we hope will become this new digital infrastructure around scientific research, there are several ways to monetize [it] that stem from mining the vast array of data we will have access to, such as providing information to pharmaceutical companies about trending areas or breakthroughs in science that haven’t yet hit their radar,“ he ([sc. Richard Proce] said. „Another example is that we’ll be able to help universities and R&D companies source the top scientists in specific areas, and hence monetize via recruiting avenues.“ [Shankland, 2013]

Der Status von Informationen, die sich durch geeignete Algorithmen aus einer großen Menge von persönlichen Daten herauspräparieren lassen, ist umstritten. Die Nutzungsbedingungen von academia.edu lassen dagegen keinen Zweifel darüber, dass die Ergebnisse ihrer Aufarbeitung restlos den Eigentümern des Portals gehören. Die Bereitstellung von Inhalten auf der Webseite kommt, so legen es die „Terms of Use“ fest, der Übertragung einer weltweiten, unkündbaren, kostenfreien Lizenz an academia.edu gleich.

Diese Lizenz erlaubt es dem Empfänger, die Inhalte in jeder erdenklichen Form zu nutzen: „to use, view, copy, adapt, modify, distribute, license, sell, transfer, publicly display, publicly perform, transmit, stream, broadcast and otherwise exploit such Member Content ...“¹¹ Unter die Generalklausel „and otherwise exploit“ fallen sicherlich jene Auswertungen, die den kommerziellen Kern des

10

<http://jurnsearch.wordpress.com/2011/12/15/academic-search-attracts-4-5-venture-capital-money/>

11 <http://www.academia.edu/terms> 4.1.2014

Unternehmens ausmachen: die Ergebnisse des „data mining“ zur Extraktion zusätzlicher Information.

Im Unterschied zur unlängst bekannt gewordenen Ausspähung der Kommunikation im Internet durch die National Security Agency der Vereinigten Staaten haben die Benutzerinnen der Plattform das Einverständnis zur Nachbearbeitung ihrer Mitteilungen vorneweg gegeben. Das Vorgehen auf academia.edu ist darum nicht illegitim. Ob allerdings die Auswertung der Inhalte, die von Benutzerinnen kommen, für die geschäftlichen Interessen eines start-ups unbedenklich ist, sei dahingestellt. Die Angaben der Webseite zum Umgang mit privaten Daten sind schwammig formuliert.¹²

academia.edu ist kein Einzelfall. Die Geschäftsbedingungen für Apples iTunes Store enthalten ähnliche Passagen. Speziell bemerkenswert ist, dass iTunes sich das Recht nimmt, eingestellte Materialien *zu verändern*.¹³

Sie räumen iTunes hiermit eine örtlich und zeitlich unbeschränkte, kostenlose und einfache Lizenz ein, solche Materialien als Teil des Dienstes und im Zusammenhang mit den iTunes Produkten zu nutzen, ohne Gegenleistung oder Verpflichtungen Ihnen gegenüber. iTunes behält sich das Recht vor, jederzeit nach seinem eigenen Ermessen und ohne Benachrichtigung und Haftung Ihnen gegenüber Materialien nicht einzustellen oder zu veröffentlichen und Materialien zu entfernen oder zu verändern.

Eine Studie über Apples legistische Strategie bemerkt, dass sich diese Regelungen auf vertragsrechtliche Prinzipien berufen, denen liberalere Grundsätze im Bereich des Copyright entgegenstehen. [Reder, 2008, S. 200f] Die juristische Bewertung der hier angesprochenen Entwicklung steht auf einem anderen Blatt. Das Augenmerk galt technischen Entwicklungen, die das Äquilibrium zwischen der Verfügung über Informationen, welche die eigene Person betreffen, und ihrer

12 „We may combine your Personal Information with Non-Identifying Information and aggregate it with information collected from other Academia.edu Users (defined below) to attempt to provide you with a better experience, to improve the quality and value of the Academia.edu Service and to analyze and understand how our Site and Academia.edu Service are used.“
<https://www.academia.edu/privacy> 4.1.2014

13 Die juristischen Details sind für jedes Land separat formuliert: www.apple.com/legal/internet-services/itunes/ww. Zitiert wird die Version für Österreich.

Verwertung in Management- und Geschäftsprozessen durcheinanderbringen.

Zu Beginn wurde darauf hingewiesen, dass sich vielfältige Statistiken Informationen zu Nutze machen, die das soziale Leben generiert. Darin alleine kann das aktuelle Problem nicht liegen. Spezifisch auf den Wissenschaftsbetrieb bezogen, lässt es sich so präzisieren: bibliometrische Verfahren haben durch die Einrichtung globaler, öffentlicher Datenkommunikation enorm an Effektivität und Einfluss gewonnen. Sie greifen durch Rankings und Erfolgsnachweise zunehmend in die Gestaltung jener Abläufe ein, welche sie darstellen sollen. Dazu leisten verordnete Bilanzen des „intellektuellen Kapitals“ einen wichtigen Beitrag. [Würtz, 2009] Weder die Objektivität, noch die Reproduzierbarkeit, noch die sachliche Adäquatheit solcher Qualitätserhebungen sind gesichert.¹⁴

Der kritische Punkt liegt dort, wo sich herausstellt, dass die epochalen Erleichterungen des Informationstransfers mit erheblichen Grenzverletzungen in bislang separierten Revieren einhergehen. Ethik bietet eine Demarkierung differenzierter Anspruchsfelder, basierend auf einem Konsens über gemeinsame Verhaltensstandards. Eine Auswirkung von „big data“ ist die Unterhöhlung der relativen Autonomie des traditionellen Wissenschaftsbetriebs, inklusive der Wertvorstellungen, die damit verbunden waren. Über die Vor- und Nachteile dieser Traditionen kann man geteilter Meinung sein. Darüber, dass sich die Beteiligten die Situation deutlich vor Augen führen müssen, um ihr Verhalten mit den dargestellten Herausforderungen konfrontieren zu können, sollte es keine Diskussion geben.

Literatur

[Anderson, 2013] Anderson, P. (2013). What's wrong with google scholar for „Systematic“ reviews.

14 Einen Überblick bietet [Lenzen et al., 2012]. „Wenn wir also zusammenfassend fragen, ob Wissenschaftsrankings ihre Zwecke erfüllen, dann ist dieses zu verneinen, weil sie nicht objektiv, nicht reliabel und nicht valide sind, ihre Skalierungsformen sind fehlerhaft und irreführend. Hinzu tritt, dass internationale Rankings oftmals nur mit sogenannten »Proxies« arbeiten, also mit Stellvertreterdaten für Qualität, dass sie sich im Wesentlichen auf ein Prozent sämtlicher Universitäten beschränken und dass im Übrigen ein Verzerrungseffekt dann eintritt, wenn englischsprachige Datenbanken als Quelle herangezogen werden und nicht auch nationalsprachliche andere.“ S. 18

<http://etechlib.wordpress.com/2013/01/23/whats-wrong-with-google-scholar-for-systematic-reviews/>. 4.1.2014

[Bar-Ilan, 2007] Bar-Ilan, J. (2007). Manipulating search engine algorithms: the case of Google. *Journal of Information, Communication & Ethics in Society*, 5(2/3):155–166.

[Bénilde, 2013] Bénilde, M. (2013). La traque méthodique de l'internaute révolutionne la publicité. *Le Monde diplomatique*, pages 22–23.

[Benkler, 2006] Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

[Bowen, 2014] Bowen, A. (2014). Is google scholar a useful source?
<http://libguides.csuchico.edu/content.php?pid=524292&sid=4312897>. 4.1.2014.

[Castells, 2011] Castells, M. (2011). *The rise of the network society: The information age: Economy, society, and culture*, volume 1. Wiley. com.

[Cerdá et al., 2009] Cerdá, J. H. C., Nieto, E. M., and Campos, M. L. (2009). What's wrong with citation counts? *D-Lib Magazine*, 15(3/4):1082–9873.

[Delgado López-Cózar et al., 2012] Delgado López-Cózar, E., Robinson-García, N., and Torres-Salinas, D. (2012). Manipulating google scholar citations and google scholar metrics: simple, easy and tempting. *arXiv:1212.0638S*.

[Dilger and Harry Müller, 2012] Dilger, A. and Harry Müller, B. (2012). Ein Forschungsleistungsranking auf der Grundlage von Google Scholar. *Zeitschrift für Betriebswirtschaft*, 82(10):1089–1105.

[Gray, 2012] Gray, J. E. e. (2012). Scholarish: Google scholar and its value to the sciences. *Issues in Science and Technology Librarianship*.

[Harzing and van der Wal, 2008] Harzing, A. and van der Wal, R. (2008). Google scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, 8:61–73. 00002.

[Harzing and Van der Wal, 2007] Harzing, A.-W. and Van der Wal, R. (2007). Google scholar: the democratization of citation analysis. *Ethics in Science and environmental Politics*, 8(1):61–73.

[Jacsó, 2005] Jacsó, P. (2005). Google scholar: the pros and the cons. *Online information review*, 29(2):208–214.

[Jacsó, 2009] Jacsó, P. (2009). Google scholar's ghost authors. *Library Journal*, 134(18):26–27.

- [Kraft, 2009] Kraft, M. (2009). The horror of google scholar. <http://kraftylibrarian.com/?p=219>. 4.1.2014.
- [Lenzen et al., 2012] Lenzen, D. et al. (2012). Ranking, Rating – Steuerung und Motivation. Erfahrungen und Befunde zum Forschungsrating aus Sicht einer Universitätsleitung. *Erziehungswissenschaft*, 23(45):13–25.
- [Mikki, 2009] Mikki, S. (2009). Google scholar compared to web of science. a literature review. *Nordic Journal of Information Literacy in Higher Education*, 1(1).
- [Noruzi, 2007] Noruzi, A. (2007). Google scholar: the new generation of citation indexes. *Libri. International Journal of Libraries and Information Services*, 55(4):170–180.
- [Reder, 2008] Reder, M. E. K. (2008). Case study of apple, inc. for business law students: How apple’s business model controls digital content through legal and technological means. SSRN Scholarly Paper ID 1085702, Social Science Research Network, Rochester, NY.
- [Shankland, 2013] Shankland, S. (2013). Academia.edu raises funds to build a facebook for scientists textbar cutting edge. http://news.cnet.com/8301-11386_3-57604722-76/academia.edu-raises-funds-to-build-a-facebook-for-scientists/ 4.1.2014.
- [Tene, 2008] Tene, O. (2008). What google knows: Privacy and internet search engines.
- [Würtz, 2009] Würtz, M.-O. (2009). *Einführung einer Wissensbilanz in einem forschungsorientierten Industrieunternehmen*. GRIN Verlag.